In situ sequencing for RNA analysis in preserved tissue and cells

Rongqin Ke[1,2,5], Marco Mignardi[1,2,5], Alexandra Pacureanu[3], Jessica Svedlund[1], Johan Botling[2], Carolina Wählby[3,4] & Mats Nilsson[1,2]

Tissue gene expression profiling is performed on homogenates or on populations of isolated single cells to resolve molecular states of different cell types. In both approaches, histological context is lost. We have developed an *in situ* sequencing method for parallel targeted analysis of short RNA fragments in morphologically preserved cells and tissue. We demonstrate *in situ* sequencing of point mutations and multiplexed gene expression profiling in human breast cancer tissue sections.

Most organism-level functions are executed by the concerted action of different cell types. The identity and function of each cell is defined by its gene expression program, which in turn is governed by the cell's path of differentiation and by external stimulation from surrounding cells and extracellular matrix. Attempts to study organ function by measuring gene expression in extracts after tissue homogenization will be flawed because the average gene expression profile across the tissue cannot be used to deduce the transcript-level molecular state of the different cell types in the tissue[1]. After decades of research on cell cultures, we need to move our focus to understand the interplay between distinct cell types in complex organ tissues. To that end, new technological approaches are required.

The application of next-generation sequencing technology to RNA sequencing has provided a more comprehensive view of the RNA content of cells than previous techniques[2]. However, the currently available technologies for RNA sequencing are based on purified nucleic acids extracted from their natural context, and such an extraction limits the possibility of connecting the sequence with spatial information, which is important when analyzing tissues consisting of mixed populations of different cell types. This limitation can be addressed by isolating individual cells from tissue sections through laser-capture microdissection, FACS or a mouth-controlled pipetting system and then analyzing nucleic acid content and sequences[3–8]. However, these techniques are low in throughput in terms of number of cells analyzed or are limited in spatial resolution. Also, because of sampling bias, the collected cells might not reflect the nature of the true cell compartment targeted for expression profiling. Thus, sequencing single molecules directly in the tissue environment is a major goal for single-cell analysis technology.

Here we show that sequencing chemistry can be applied *in situ* for analysis of up to four-base-pair fragments in single mRNA molecules in the unperturbed context of fixed cells and tissues. Our method is based on padlock probing, rolling-circle amplification (RCA) and sequencing-by-ligation chemistry[9–13] (**Fig. 1**). RCA in combination with padlock-probe circularization reactions has been used to produce clonally amplified rolling-circle products (RCPs) at high density in cells and tissue sections at micrometer spatial resolution, enabling detection and genotyping of individual mRNA molecules *in situ*[11]. Here we subjected RCPs to sequencing by ligation to read short segments of RNA or sequence tags. We have developed two targeted approaches (gap and barcode targeted; **Fig. 1**) to analyze selections of transcripts in tissue. Owing to the high density of transcripts in the cells, we applied targeted sequencing to avoid crowding of RCPs, which would make base-calling difficult. We have created a fully automated image analysis pipeline for base-calling in the open-source software CellProfiler calling ImageJ plug-ins for image alignment[14,15], in which images from consecutive sequencing cycles are automatically aligned and fluorescence intensities are measured at every position corresponding to a sequencing substrate (**Supplementary Fig. 1** and Online Methods).

To test the feasibility of *in situ* sequencing of mRNA fragments, we sequenced a stretch of four different nucleotides in the human β-actin transcript (*ACTB*) using gap-targeted sequencing. We successfully called all of these four bases with an average accuracy of 98.6% (19 errors in 1,316 base calls from ten fibroblast cells) (**Supplementary Fig. 2**). High spatial correlation of signals from different imaging cycles indicated that the RCPs, which constitute the sequencing substrates, did not move measurably between imaging cycles, which otherwise would have made base-calling difficult. There is no chemical fixation step in the protocol after the generation of RCPs, so the nature of this bond to the cellular interior is unknown. There was virtually no loss of sequencing substrates during repeated cycles, but there was a tendency for substrates producing weak signal to fall below the detection threshold owing to increased background after repeated sequencing cycles (**Supplementary Fig. 3**).

To verify the specificity of the sequencing, we used the gap-targeted approach to sequence a different four-base-long sequence motif of *ACTB* in cocultures of human and mouse fibroblasts

[1]Science for Life Laboratory, Department of Biochemistry and Biophysics, Stockholm University, Stockholm, Sweden. [2]Department of Immunology, Genetics, and Pathology, the Rudbeck Laboratory, Uppsala University, Uppsala, Sweden. [3]Science for Life Laboratory, Centre for Image Analysis, Uppsala University, Uppsala, Sweden. [4]Imaging Platform, Broad Institute of Harvard and Massachusetts Institute of Technology, Cambridge, Massachusetts, USA. [5]These authors contributed equally to this work. Correspondence should be addressed to M.N. (mats.nilsson@scilifelab.se) or C.W. (carolina@broadinstitute.org).

RECEIVED 20 APRIL; ACCEPTED 20 JUNE; PUBLISHED ONLINE 14 JULY 2013; DOI:10.1038/NMETH.2563

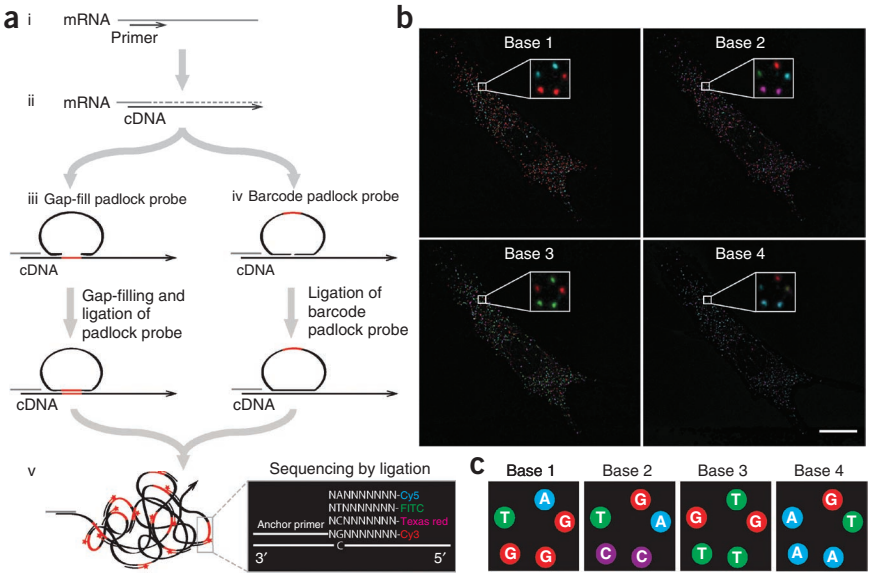

Figure 1 | Procedure for targeted *in situ* sequencing. (**a**) mRNA is copied to cDNA by reverse transcription (**i**), which is followed by degradation of the mRNA strand by RNase H (**ii**). A padlock probe is hybridized to the cDNA strand. (**iii**) For gap-targeted sequencing, the padlock probe binds to the cDNA with a gap between the probe ends over the bases that are targeted for sequencing by ligation. This gap is filled by DNA polymerization and DNA ligation to create a DNA circle. (**iv**) For barcode-targeted sequencing, DNA circularization of a padlock probe carrying a barcode sequence is carried out by ligation only. (**v**) The DNA circle is amplified by target-primed RCA generating an RCP that is subjected to sequencing by ligation. An anchor primer is hybridized next to the targeted sequence (red fragment) before the ligation of interrogation probes, which consists of four libraries of 9-mers, with eight random positions (N) and one fixed position (A, C, G or T); each library is labeled with one of four fluorescent dyes. The interrogation probe with best match at the fixed position is incorporated by ligation, along with the corresponding fluorophore. (**b**) The sample is imaged, and each RCP displays the color corresponding to the matched base. The interrogation probe is washed away before the application of interrogation probes for the next base. These steps of ligation, imaging and washing are iterated until the desired number of bases has been read. Enlarged views of five RCPs are shown for each sequencing cycle. The illustration is based on actual data. Scale bar, 50 µm. (**c**) Base-calling is done by image analysis recording the fluorescence staining patterns across sequencing cycles.

(BJ-hTERT and MEF, respectively). The sequence is identical in the two species except for a single-nucleotide variation at the second position from the sequencing primer. All reads corresponded to the two expected sequences, and the majority of reads were mapped to the expected cell type (208 out of 215 reads from five cells; **Supplementary Fig. 4**). Mapping to the correct cell may be improved by adding a cytoplasmic stain to guide the automated cell segmentation.

To test the sensitivity of finding a rare cell type in a large background of other cells, we performed *in situ* sequencing of codon 12 of the *KRAS* transcript in cell cultures with different spike-in ratios (1:10, 1:100 and 1:1,000) of two cell lines (*KRAS* mutant A549 spiked in wild-type ONCO-DG1 background). We were able to detect the mutations on a per-cell basis, in ratios that corresponded well with the spiked-in proportions of the two populations (**Supplementary Table 1**). For this experiment, we applied automated scanning (Online Methods) to generate data from thousands of cells.

To prove that sequencing of transcript fragments could be performed directly in tissue sections, we sequenced four-base-pair motifs of *ACTB* and *HER2* (*ERBB2*) in a section of a *HER2* transcript–positive ('*HER2*-positive') human fresh-frozen breast cancer tissue using gap-targeted sequencing. We generated 298 *ACTB* reads and 27 *HER2* reads, located in different parts of the tissue section, in an image covering a tissue area of 222 × 166 µm^2 (**Fig. 2** and **Supplementary Fig. 5**). The *HER2* transcript was detected only in the area of the section corresponding to cancerous tissue (according to morphological analysis). Four reads did not match the expected two sequences (with one mismatch per read), a result corresponding to a base-calling accuracy of 99.7%.

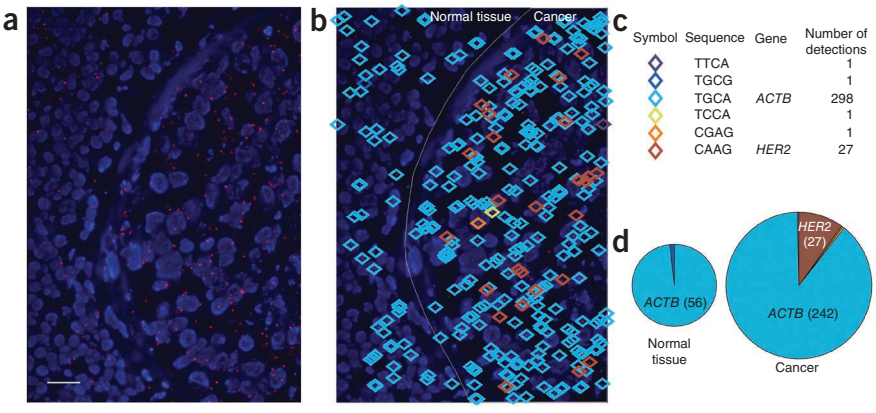

Figure 2 | *In situ* sequencing of fragments of *ACTB* and *HER2* mRNA in breast cancer tissue. (**a**) Raw data showing the location of sequences called from a fresh-frozen breast cancer tissue section (blue, DAPI; red, general stain of sequence common to all probes). Scale bar, 25 µm. (**b**,**c**) Each diamond symbol represents a decoded sequence, color coded as shown. The white line was manually drawn to separate cancer cells from adjacent nonmalignant stroma. (**d**) The relative frequency of each sequence is quantified in normal and cancer tissue and is represented by a pie chart (the number in parentheses is the number of occurrences, and the total area is proportional to the total number of RCPs). The two most abundant sequences are from *ACTB* (light blue) and *HER2* (brown) transcripts. Note that the other unexpected sequences differ by a single nucleotide and occur only once each.

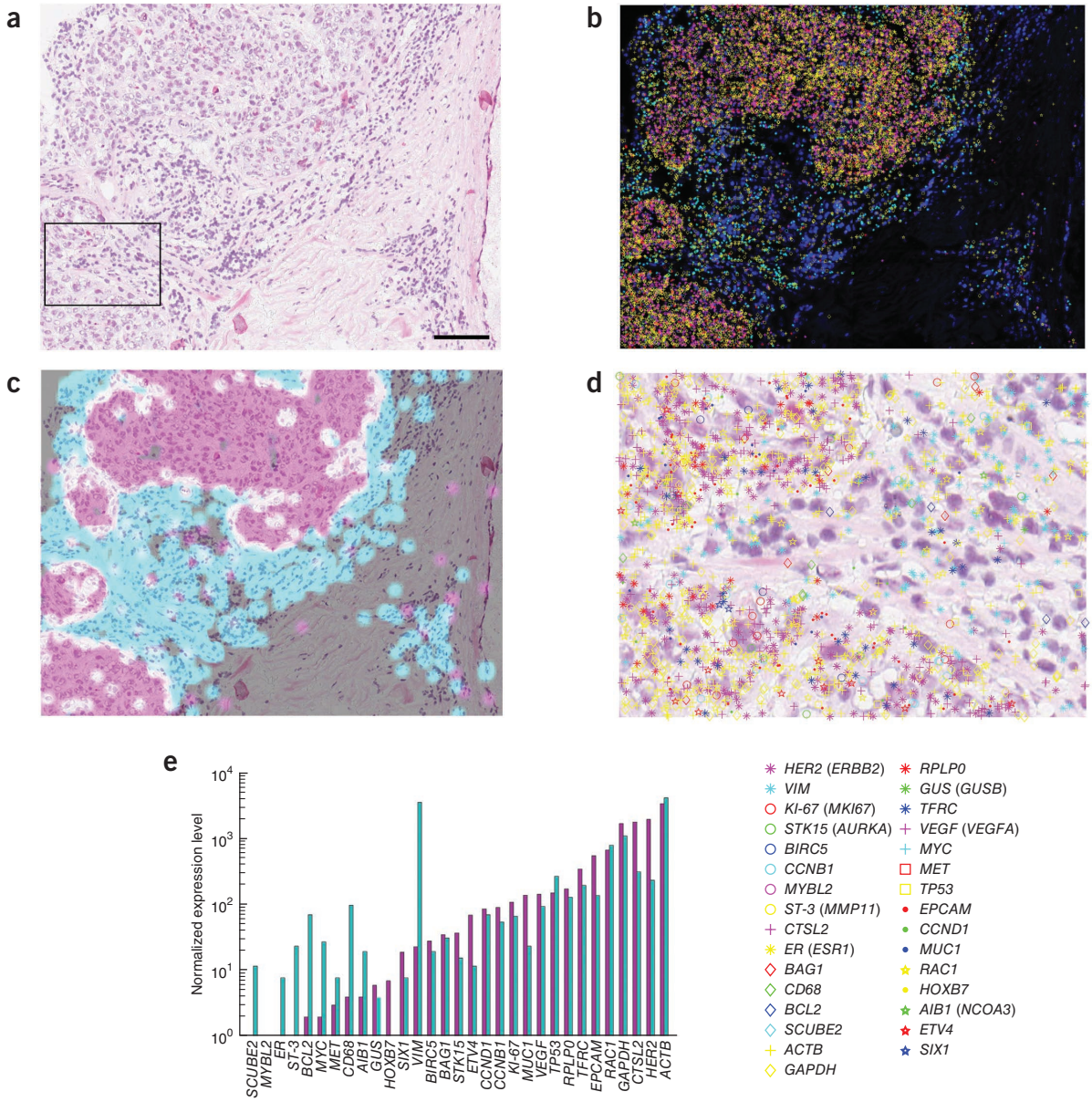

Figure 3 | Gene expression profiling on an *ER*-negative breast cancer tissue section by barcode-targeted *in situ* sequencing. (**a**) Part of a hematoxylin-and-eosin (H&E)-stained fresh-frozen breast cancer tissue section that was subjected to sequencing. Scale bar, 100 µm. (**b**) Localization of each of the 31 detected barcodes is shown as a symbol on top of a fluorescence microscopy image showing the nuclei (blue). Each symbol represents a barcode sequence that corresponds to a specific transcript. (**c**) Map showing the local density of detected *HER2* (magenta) and *VIM* (cyan) transcripts, plotted on top of the H&E stain image. White indicates coexpression of *HER2* and *VIM*. Note that the molecular staining pattern aligns to the histological staining pattern. (**d**) Spatial location of 31 different transcripts in the boxed area in **a**. (**e**) Expression profiling of the same 31 transcripts in the *HER2*-positive region (in magenta) and *VIM*-positive region (cyan). Transcript counts in the *VIM*-positive region were normalized to those in the *HER2*-positive region.

The current sequencing read-length capability of the system limits the possibility of finding distinct tag sequences in large sets of transcripts and may thereby limit highly multiplex *in situ* detection of mRNA and studies of more complex sequence and splice variants. However, four bases of sequence information are enough to decode $4^4 = 256$ different sequence barcodes. To demonstrate this, we designed a padlock-probe targeting human *ACTB* that included four degenerated nucleotides (NNNN) that upon sequencing should generate up to 256 different sequencing reads. We combined this probe with a sequencing control probe for human *GAPDH* carrying one unique barcode. Sequencing 35 human fibroblast cells, we retrieved 4,412 reads for *ACTB* from 232 barcodes and 2,949 reads for *GAPDH*, with a base-calling accuracy for *GAPDH* of 99.2% (**Supplementary Fig. 6**).

We applied the barcode-targeted approach for highly multiplexed *in situ* expression profiling in individual tissue sections. We designed padlock probes targeting 39 transcripts, including 21 transcripts used in a breast cancer prognostic expression panel (OncoType DX)[16], to study the transcript expression and localization in three *HER2*-positive fresh-frozen breast cancer tissue sections (slides A, B and C). We generated cDNA from total RNA *in situ* by random-decamer priming, avoiding the use of

the 39 locked nucleic acid (LNA)-modified primers specific for the target RNAs. We applied all padlock probes in one reaction and determined the expression pattern for the corresponding genes by sequencing each probe's unique four-base-long barcode *in situ*. We extracted expression data from 31 transcripts; for a transcript's frequency to be considered reliable, we required that the frequency of detection be higher than that of the most abundant unexpected barcode read (**Supplementary Fig. 7**). The detected transcripts displayed different patterns of localization across the tissues, and the number of signals varied for the different targeted transcripts. The patterns do not seem random (**Fig. 3**, **Supplementary Figs. 8–10** and **Supplementary Note 1**). Hematoxylin and eosin staining of the same tissue sections showed that *HER2* expression was localized mainly in the cancer cell compartment, whereas *VIM* was localized in infiltrating lymphocytes and other components of the stroma (**Fig. 3** and **Supplementary Figs. 8–11**). We were able to verify, using our expression profiling approach, that one of the tissue samples (slide A) was derived from an estrogen receptor (*ER* or *ESR1*)-negative tumor, whereas the other two (slides B and C) were from an *ER*-positive tumor (**Fig. 3** and **Supplementary Fig. 12**). To further evaluate how well our *in situ* expression data relate to expression profiles generated by *in vitro* approaches and to estimate the sensitivity of the *in situ* approach, we compared our *in situ* expression profiling data with published RNA-seq data[17] from one normal breast tissue, one human mammary epithelial cell line and three breast cancer–derived cell lines. The greatest similarity for our *VIM*-positive *in situ* expression profile was with the RNA-seq data for normal breast tissue (Pearson correlation, $r = 0.66$), and the highest correlation for our *HER2*-positive *in situ* expression profile was with the breast cancer–derived cell line Bt474 ($r = 0.76$) (**Supplementary Table 2** and **Supplementary Fig. 13**). The expression measured by RNA-seq and by *in situ* sequencing correlated well over a broad range of expression in both comparisons (**Supplementary Table 3**).

To determine the maximum number of reads that we could analyze per area unit, we investigated the data from the *HER2*-positive area shown in **Figure 3**, which has a high density of reads. The *HER2*-positive area comprised about 450 cells covering an area of 0.16 mm^2. In this area we observed 11,423 reads (on average, 25 reads per cell) from RCA products that had an average diameter of 1.0 µm. Combined, they covered 5.5% of the area. By simulation, this area occupancy can be increased up to 20% before loss of signals due to overlap becomes substantial (**Supplementary Fig. 14**). Thus, from these numbers we estimated that the maximum information content with the current RCA size was about 270,000 reads per mm^2, which corresponds to around 90 reads per cell at a cell density similar to that of the *HER2*-positive tissue area we analyzed. Theoretical boundaries of the approach are further discussed in **Supplementary Note 2**.

Gene expression profiling by *in situ* sequencing of barcodes reveals tissue heterogeneity at the molecular level. With enough transcripts detected, it might be possible to identify different types of cells on the basis of their expression profile with a greater resolution than those of current state-of-the-art methods. Unsupervised clustering of the local spatial expression profiles resulted in clear *HER2*-positive and *HER2*-negative tissue regions (**Supplementary Fig. 15**), indicating that this approach could be used to characterize tissue compartments with distinct expression patterns without a priori information.

Our method allows hypothesis-driven targeted analyses of multiple RNA sequences and variants thereof in preserved cells and tissue. With this technology it will be possible to decompose expression profiles obtained from tissue homogenates and to assign expression profiles to the underlying cellular components of a tissue. It provides new possibilities for studying complex biological events such as cell differentiation in heterogeneous populations of cells in their natural context. In conclusion, we have sequenced small RNA fragments *in situ* for the first time, to our knowledge, opening up new prospects for the use of sequencing technology in basic research and clinical diagnostics.

METHODS

Methods and any associated references are available in the online version of the paper.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

We thank J. Lee, G.M. Church and F. Pontén for valuable discussion about this work. We thank M. Dahlberg for helping extracting RNA-seq data from publications. The research reported in this paper was funded by the Swedish Research Council, VINNOVA project "Companion diagnostic initiative," the European Community's 7th Framework Program (FP7/2007-2013) under grant agreement nos. 259796 (DiaTools) and 201418 (READNA), the Science for Life Laboratory, Stockholm and Uppsala, and the Innovative Medicines Initiative Joint Undertaking under grant agreement no. 115234 (OncoTrack).

AUTHOR CONTRIBUTIONS

R.K. and M.M. designed and performed the experiments. J.B. provided tissue sections and pathology examination of the tissue. C.W. designed the image analysis pipelines and performed the image analysis together with A.P. J.S. performed the correlation between *in situ* sequencing and RNA-seq data. R.K., M.M., C.W. and M.N. wrote the manuscript. All authors commented on and revised the manuscript. M.N. conceived the idea and supervised the project.

COMPETING FINANCIAL INTERESTS

The authors declare competing financial interests: details are available in the online version of the paper.

1. Levsky, J.M. & Singer, R.H. *Trends Cell Biol.* **13**, 4–6 (2003).
2. Wang, Z., Gerstein, M. & Snyder, M. *Nat. Rev. Genet.* **10**, 57–63 (2009).
3. Bonner, R.F. *et al. Science* **278**, 1481–1483 (1997).
4. Dalerba, P. *et al. Nat. Biotechnol.* **29**, 1120–1127 (2011).
5. Navin, N. *et al. Nature* **472**, 90–94 (2011).
6. Tang, F. *et al. Nat. Methods* **6**, 377–382 (2009).
7. Hou, Y. *et al. Cell* **148**, 873–885 (2012).
8. Xu, X. *et al. Cell* **148**, 886–895 (2012).
9. Nilsson, M. *et al. Science* **265**, 2085–2088 (1994).
10. Banér, J., Nilsson, M., Mendel-Hartvig, M. & Landegren, U. *Nucleic Acids Res.* **26**, 5073–5078 (1998).
11. Larsson, C., Grundberg, I., Soderberg, O. & Nilsson, M. *Nat. Methods* **7**, 395–397 (2010).
12. Shendure, J. *et al. Science* **309**, 1728–1732 (2005).
13. Drmanac, R. *et al. Science* **327**, 78–81 (2010).
14. Kamentsky, L. *et al. Bioinformatics* **27**, 1179–1180 (2011).
15. Thévenaz, P., Ruttimann, U.E. & Unser, M. *IEEE Trans. Image Process.* **7**, 27–41 (1998).
16. Sparano, J.A. & Paik, S. *J. Clin. Oncol.* **26**, 721–728 (2008).
17. Wang, E.T. *et al. Nature* **456**, 470–476 (2008).

ONLINE METHODS

Cell culture and sample preparation. The cell lines BJ-hTERT, MEF, A-549 and GM-8402 were cultured in DMEM (Gibco) without phenol red and L-glutamine, supplemented with 10% FBS (Sigma), 2 mM L-glutamine (Sigma) and 1× PEST (Sigma). The cell line ONCO DG-1 was cultured in RPMI 1640 (Sigma) without phenol red and L-glutamine, supplemented with 10% FBS (Sigma), 2 mM L-glutamine (Sigma) and 1× PEST (Sigma). All cell lines were incubated at 37 °C, 5% CO_2.

To prepare cell samples, we treated confluent cell lines with 0.25% (w/v) trypsin-EDTA (Sigma) and resuspended them in culturing medium. Resuspended cells were then seeded on Superfrost Plus slides (Thermo) placed in a 150 mm × 25 mm Petri dish (Corning), and culturing medium was added to a final volume of 25 ml. Three milliliters of resuspended cells were used to seed five slides. Cells were incubated in the same previous conditions 12–24 h before fixation. Slides with cocultured cell lines were prepared as described above with a mixture of different cell lines. Fixation was performed in 3% (w/v) paraformaldehyde (Sigma) in DEPC-treated PBS for 30 min at RT after removal of the culturing medium and two washes in PBS. After fixation, slides were washed twice in DEPC-treated PBS and dehydrated in an ethanol series of 70%, 85% and 100% for 5 min each. All of the following reactions were performed in Secure-Seal hybridization chambers (Invitrogen).

Tissue sections. Frozen sections (4 µm) from fully 'anonymized' human fresh-frozen breast cancer tissues were obtained from the biobank at the Department of Pathology, Uppsala University Hospital, in accordance with the Swedish Biobank Legislation. Tissue samples were stored at −80 °C until fixation. Fixation was performed in 3% (w/v) paraformaldehyde (Sigma) in DEPC-treated PBS with 0.05% Tween-20 (Sigma) (DEPC-PBS-T) for 45 min at RT and was followed by two washes in DEPC-PBS-T. The tissue was permeabilized in 2 mg/ml pepsin (Sigma) in 0.1 M HCl at 37 °C for 90 s. After permeabilization, slides were washed twice in DEPC-treated PBS and dehydrated in an ethanol series. All of the following reactions were performed in Secure-Seal hybridization chambers (Invitrogen).

***In situ* reverse transcription.** Both cells and tissue samples were first rinsed with DEPC-PBS-T. For the cell samples, this was followed by incubation in 0.1 M HCl (in DEPC H_20) for 5 min and two washes with DEPC-PBS-T. For tissue, no further treatments were performed after the steps described above. Then the reversed transcription mix, containing 1 µM of LNA-modified cDNA primer or 5 µM of unmodified random decamers (all oligonucleotides sequences are listed in **Supplementary Tables 4–6**), 20 U/µl of RevertAid H Minus M-MuLV reverse transcriptase (Fermentas), 500 µM dNTPs (Fermentas), 0.2 µg/µl BSA (NEB) and 1 U/µl RiboLock RNase Inhibitor (Fermentas) in the M-MuLV reaction buffer, was added on the slides. The incubation was carried out for 3 h at 37 °C for LNA primers and overnight for random decamers. Slides were washed twice with PBS-T, which was followed by a postfixation step in 3% (w/v) paraformaldehyde in DEPC-PBS for 10 min at room temperature (30 min for tissues). After postfixation, the samples were washed twice in DEPC-PBS-T.

Gap-fill padlock probing. All padlock probes were phosphorylated before use. A mixture containing 2 µM of padlock probes, 1× PNK buffer A (Fermentas), 1 mM ATP, 0.1 U/µl T4 polynucleotide kinase was incubated at 37 °C for 30 min and at 60 °C for 20 min. The phosphorylated padlock probes can be stored at −20 °C until used. After reverse transcription, a mix that performs the degradation of RNA, hybridization of the padlock probe, gap-filling (copy) of the target sequence for sequencing, and ligation to form a complete DNA circle was added. The mix contained 1× Ampligase buffer (20 mM Tris-HCl, pH 8.3, 25 mM KCl, 10 mM $MgCl_2$, 0.5 mM NAD and 0.01% Triton X-100), 100 nM of each padlock probe, 50 µM dNTPs, 0.2 U/µl Stoffel fragment (Applied Biosystems), 0.5 U/µl Ampligase, 0.4 U/µl RNase H (Fermentas), 1 U/µl RiboLock RNase Inhibitor, 50 mM KCl and 20% formamide. The incubation was carried out at 37 °C for 30 min and then held at 45 °C for 45 min. Then the slide was washed twice with 1× DEPC-PBS-T.

Barcode padlock probing. For gene expression profiling in tissue, a total of 39 padlock probes (phosphorylated as describe above) were applied. Fifty microliters of ligation mix containing 1× Ampligase buffer, 100 nM of each padlock probe, 0.5 U/µl Ampligase, 0.4 U/µl RNase H (Fermentas), 1 U/µl RiboLock RNase Inhibitor, 50 mM KCl and 20% formamide was added to each reaction chamber. The incubation was carried out at 37 °C for 30 min and then held at 45 °C for 45 min. Then the slide was washed with 1× DEPC-PBS-T twice.

RCA and RCA products detection. An RCA mix containing 1 U/µl phi29 polymerase (Fermentas), 1× phi29 polymerase buffer, 0.25 mM dNTPs, 0.2 µg/µl BSA and 5% glycerol in DEPC H_2O was added to the reaction chamber and incubated for 2 h to overnight to carry out the RCA. After the incubation, the slide was washed three times in DEPC-PBS-T. Finally, 100 nM of each corresponding detection probe (uracil-containing detection probes were used in KRAS sequencing as well as tissue sequencing) in 2× SSC and 20% formamide was applied to the slide and incubated at 37 °C for 30 min. Excess amounts of detection probes were then eliminated by washing three times with DEPC-PBS-T. After the washing, the secure seals were removed. The slides were then dehydrated through an ethanol series. The slides were mounted in Vectashield mounting medium (Vector) that contains 100 ng/ml of DAPI (4′,6-diamidino-2-phenylindole) for counterstaining the nuclei and were analyzed using an AxioplanII epifluorescence microscope (Zeiss).

Sequencing by ligation. Before the sequencing was performed, the detection probes were stripped off. The slides were first incubated through an ethanol series to remove the mounting medium and dried at room temperature. For detection probes without uracils, the samples were first washed with DEPC-PBS-T and then incubated three times with 65% formamide for 30 s, which was followed by washing twice with DEPC-PBS-T. Detection probes that contained uracils were first treated with UNG treating buffer (1× phi29 polymerase buffer (Fermentas), 0.2 µg/µl BSA, 0.02 U/µl UNG (Fermentas)) for 10 min and washed twice with DEPC-PBS-T before the formamide incubation.

A mix containing 500 nM of corresponding anchor primers in 2× SSC and 20% formamide were then added to the sample and incubated at RT for 30 min, and the incubation was followed by two brief washes with DEPC-PBS-T. A ligation mix containing

each interrogation probe, 1× T4 ligase buffer (Fermentas), 1 mM ATP (Fermentas) and 0.1 U/μl of T4 ligase (Fermentas) was applied to the samples and incubated for 30 min at RT. The unligated probes were washed away by 3× 1 min incubation with DEPC-PBS-T. The slides were mounted in Vectashield mounting medium containing 100 ng/ml of DAPI for counterstaining the nuclei. The concentration of each interrogation probe was 100 nM for cell cultures and 500 nM for tissue sections.

After imaging, the slides were prepared for the next sequencing cycle by UNG treatment buffer as described above followed by repeating the hybridization, ligation and imaging processes. To sequence each base, we applied the same procedures. For evaluation of loss of sequencing substrates (RCA products), cell cytoplasm was stained after the last sequencing cycle with Alexa Fluor 488 phalloidin (Invitrogen) in PBS at a final concentration of 0.15 μM, and the cells were incubated for 10 min at RT.

Image acquisition and analysis. Images for all experiments, except for the spike-in experiment, were acquired using an AxioplanII epifluorescence microscope. To capture signals from RCPs located in slightly different focal planes, we captured a series of images at different focal depths. The stacks of images were thereafter merged to a single image using the maximum-intensity projection (MIP) in the Zeiss AxioVision software. Exposure times for all the experiments are listed in **Supplementary Table 7**. The images for gene expression profiling were generated by stitching 16 images (4 × 4) obtained by using a 40× objective. The automated image acquisition for the spike-in experiment was performed using a Zeiss Imager Z2 microscope with the scanning mode. A 10% overlap was set between two neighboring images. Stacks of images were captured at different focal depths and combined to a single image by MIP using the Zen software. The resulting images were then automatically stitched together into a single image containing the entire scanned area. Finally, the stitched image was used for further image analysis. For large fields of view (several mm^2), this can take hours.

The fully automated sequence decoding was performed as follows. First, images were cropped to a fixed size to remove edge effects resulting from the MIP. Cell nuclei were defined by automated thresholding and separated from each other on the basis of shape descriptors (**Supplementary Fig. 1**). Each cell's cytoplasm was defined using the nucleus as a seed in seeded watershed segmentation on an image constructed by merging all image channels from the first hybridization step and relying on sufficient cytoplasmic autofluorescence to distinguish it from the background. In cases where background fluorescence was low and such an approach failed, the cytoplasms were defined as all pixels within a fixed distance from a nucleus. The image of the general stain from the first hybridization step was filtered by a top-hat filter to enhance RCPs. Touching RCPs were separated by watershed segmentation, and each RCP was given a unique label. Prior to extracting the fluorescence intensity from each of the signals representing A, C, T and G (enhanced by top-hat filtering) using the labels from the previous step as a template, we aligned the images. The optimal transformation between a merged image of all signals (A + C + T + G) and the general stain from the first hybridization step was found using a pyramid approach to subpixel registration based on intensity[15]. All steps were performed using CellProfiler (v.2.0 r10997) calling ImageJ plugins from Fiji for image registration. A complete executable and commented CellProfiler example pipeline together with raw image data is available at http://www.cellprofiler.org/examples.shtml. All intensity information was saved to a .csv file and decoded using a script written in Matlab (also provided together with the CellProfiler pipeline). In short, for each RCP and hybridization step, the RCP was assigned the base with the highest intensity. A quality score was also extracted from each base, defined as the maximum signal (i.e., intensity of assigned letter) divided by the sum of all signals (letters) for that base. The quality of a transcript was defined as the lowest quality of all the bases in the transcript: $Q_{transcript}$ = min(hybridization step 1(max signal/all signals), …, hybridization step n(max signal/all signals)).

The quality score ranges from 0.25 to 1. A value close to 0.25 means poor quality (similar signal for all letters), whereas a value close to 1 means that the signal of the assigned letter is strong above a low background. After a typical quality threshold of 0.5–0.55 was set, the frequency of each sequence was extracted.

doi:10.1038/nmeth.2563